UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/076,194	02/14/2002	Mingkun Li	US020037	3189

24737	7590	02/01/2005

PHILIPS INTELLECTUAL PROPERTY & STANDARDS
P.O. BOX 3001
BRIARCLIFF MANOR, NY 10510

EXAMINER
KERN, MATTHEW C

ART UNIT	PAPER NUMBER
2654	

DATE MAILED: 02/01/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 10/03)

	Application No.	Applicant(s)
Office Action Summary	10/076,194	LI ET AL.
	Examiner	**Art Unit**	
	Kern Matthew	2654	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) FROM
THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____ .

2a)☐ This action is **FINAL.**   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-20* is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-20* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

       1.☐ Certified copies of the priority documents have been received.

       2.☐ Certified copies of the priority documents have been received in Application No. _____ .

       3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
    Paper No(s)/Mail Date <u>02/14/02</u>. ~~09/12/03~~ 9/12/03.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: _____ .

DETAILED ACTION

Claim Rejections - 35 USC § 112

1.   The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly
> claiming the subject matter which the applicant regards as his invention.

Claim 5 recites the limitation "plurality of image features". There is insufficient

antecedent basis for this limitation in the claim. The examiner has interpreted this to be

a reference to the "object features" in claim 1.

Claim Rejections - 35 USC § 102

2.   The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that

form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless —(a) the invention was known or used by others in this
> country, or patented or described in a printed publication in this or a foreign country, before the invention
> thereof by the applicant for a patent.

Claims 1,2,4,5,8,11, 15,16 and 17 are rejected under 35 U.S.C. 102(a) as being

anticipated by Basu et al. (US patent 6,219,640).

As per claim 1, Basu et al. teach an audio-visual (audio visual; title) system and

stored software (col. 13, lines 55-58):

-       an object detection module capable of providing a plurality of object features from

the video data (Visual and speech feature extraction, fig 3, element 22, mouth, other

facial features, col 4, lines 13-14)

- an audio processor module capable of providing a plurality of audio features from the video data (audio feature(A) extraction, element 14, fig.1, acoustic feature vectors(signals), col.4 ,lines 63-64)

- a processor coupled to the object detection and the audio segmentation modules (processor, element 602, fig.6), wherein the processor is arranged to determine a correlation between the plurality of object features and the plurality of audio features (a level of correlation between the signals, col.2, lines 35-36).

As per claim 2, Basu et al. teach a processor arranged to determine whether an animated object in the video data is associated with audio (determine the level of correlation between the signals, col.2, lines 35-36).

As per claim 4, Basu et al. teach that the animated object is a face (locate and track a face, other facial features, col 4, lines 12-13) and where the processor is arranged to determine whether the face is speaking (phonetic/visemic information from the geometry of the lip contour and its time dynamics, col. 10, lines 53-55).

As per claim 5, Basu et al. teach eigenfaces that represent global features of the face (in "Distance from Face Space" DFFS, lines, col 7. lines 32-35, feature vectors, col. 8, lines 7-8).

As per <u>claims 8,15 and 16</u>, Basu et al. teach a system implementing a method of identifying a speaking person (speaker recognition and utterance verification, title) within video data, the method comprising:

-        receiving video data including image (fig 1, element 4) and audio (figure 1, element 6) information;

-        determining a plurality of face image features from one or more faces in the video data (sub-features, hairline, chin mouth, eyes, eyebrows, col 7, lines 55-57);

-        determining a plurality of audio features related to audio information (extracts spectral features, col. 4, lines 61-63);

-        calculating a correlation between the plurality of face image features and the audio features (a level of correlation between the signals, col. 2, lines 34-35); and

-        determining the speaking person based upon the correlation (highest score identified as the speaker, col 10, lines 10-11).

As per <u>claim 11 and 17</u>, Basu et al. teach a determining step where it includes determining the speaking person based upon the one or more faces that has the largest correlation (highest combined score is identified as the speaker col 10, lines 10-11).

Claim Rejections - 35 USC § 103

3.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set
forth in section 102 of this title, if the differences between the subject matter sought to be patented and
the prior art are such that the subject matter as a whole would have been obvious at the time the
invention was made to a person having ordinary skill in the art to which said subject matter pertains.
Patentability shall not be negatived by the manner in which the invention was made.

Claim 3 is rejected under 35 USC 103(a) as being unpatentable over Basu as

applied to claim 2, in view of Nevenka (US Patent application Publication

2003/0108334).

Basu does not teach the audio features comprising two or more of the following:

Average energy, pitch, zero crossing, bandwidth, band central, roll off, low ratio,

spectral flux, or 12 MFCC components. Nevenka, however, teaches more than two

(para[0065], lines 9-11). It would have been obvious for one of ordinary skill at the time

of invention to extract and measure these acoustic features since they could provide for

a more accurate assessment when determining a person's identity.

4.   Claims 6 and 7 are rejected under 35 USC 103(a) as being unpatentable over

Basu as applied to claim 1, in view of Bradford et al.(US Patent Application Publication

2002/0103799).

As per claim 6, Basu does not teach a latent semantic indexing (LSI) module

(coupled to the processor) that preprocesses the plurality of object features and the

plurality of audio features before the correlation is performed. However, Bradford

teaches that to latent semantic indexing can be used to process both audio and text

information vectors (para. 0079, lines 8-10). It would have been obvious for one of

ordinary skill at the time of invention to have Basu's system be supplemented by the LSI

component so that a deeper level of abstraction can be achieved (para [0035], lines 10-11).

As per claim 7, Basu does not teach a latent semantic indexing module including a singular value decomposition (SVD) module. However, Bradford teaches using a SVD module (figure 2, para[0029]) to reduce term x Doc matrix to a product of three matrices. It would have been obvious for one of ordinary skill at the time of invention to have Basu's system incorporate an SVD module so that a vector space of reduced dimensionality could be produced in order to perform LSI more easily (Bradford, .

5.   Claim 9,10,12,13,14,18,19, and 20 are rejected under 35 USC 103(a) as being unpatentable over Basu as applied to claim 8, in view of Wang et al.(IEEE signal processing magazine, Nov. 2000).

As per claim 9, Basu does not teach normalizing the vectors containing the video/audio features. Wang, however, teaches normalizing these vectors (normalized correlation matrix pg 20, lines 2). It would have been further obvious to one having ordinary skill in the art at the time of invention to have Basu's system normalize the audio/video vectors with their respective information in order to better interpret the correlation, if any, that exists between the feature vectors, to see if they provide independent information, as taught by Wang (p19, col 2, lines 1-5).

As per <u>Claims 10 and 18</u>, Basu does not teach performing a singular value

decomposition on the normalized face image features and audio features. Wang,

however, teaches SVD on a normalized correlation matrix (pg 20, col 1, line 1 and col 2,

lines 4-5 (KLT-Karhunen Loeve transform)). Therefore, it would have been obvious for

one of ordinary skill at the time of invention to perform SVD on the normalized

correlation matrix as described by Wang in Basu's voice and audio speaker detection

system so that the user could determine the amount of dependence between the video

and audio features.

As per <u>Claims 12</u>, Basu does not teach a calculating step which includes forming

a matrix of the face image features and the audio features. Wang, however, teaches

combining the two in a single matrix (14 audio features, last six motion features, figure 9

and pg 20, lines 8-10). It would have been further obvious to one having skill in the art

at the time of invention to include in Basu's system the combination of both video and

audio features in a singe matrix form Wang so that the dependence among features

within the same and across different modalities could be computed, as taught by Wang

(pg 19, lines 5-8).

As per <u>Claims 13 and 19</u>, Basu does not teach performing an optimal

approximate fit using smaller matrices as compared to full rank matrices formed by the

face image features and audio features. Wang, however, teaches using SVD to allow

for dimensionality reduction (pg 10, lines 18-19). It would have been obvious for one of

ordinary skill at the time of invention to have Basu's system perform SVD for optimal approximate fit using smaller matrices in order to reduce the size of the needed eigenspace.

Claims 14 and 20 are rejected under 35 USC 103(a) as being unpatentable over Basu as applied to claim 13. Basu does not teach choosing the rank of the smaller matrices to remove noise and unrelated information from the full rank matrices. However, the examiner takes Official Notice that it is old and well-known in the art to choose the rank of the derived matrix so as to remove unrelated (and thus noisy) information from the original feature matrix. Therefore, it would have been obvious for one of ordinary skill at the time of invention to make the rank in Basu's smaller matrices such that noise and unrelated information is removed from the larger matrix, so as to get a sharper correlation between audio and video information.

Conclusion

6.   The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

Schuetze (European Patent 687,987) teaches using SVD to allow for optimal approximate fit using smaller matices.

Junqua (US patent 6,324,512) teaches a motivation for using SVD for optimal approximate fit using smaller matrices.

Yang ("Multimodal People ID for a Multimedia Meeting Browser") teaches using

SVD to reduce feature space (Karhunen-Loeve transform) of audio and video

information.

Bellegarda et al. (US patent 6,208,971) teaches a command recognition system

using SVD.

Philomin et al. (US Patent Application Publication 2003/0113002) teaches video

and audio eigen features with normalization methods.

Maali et al. (US patent 6,567,775) teach audio/video based identification system

not using a correlation matrix, but a scoring system.

Lu et al. (US patent 5,331,544) teach using eigenfaces in a marketing system.

Maes et al. (US patent 6,411,933) teach a general method of correlating biometic

data with audio information.

7.   Any inquiry concerning this communication should be directed to Mr. Matthew

Kern, whose telephone number is (703) 305-4828 or fax number (703) 305-9508. The

examiner can normally be reached Mondays-Fridays from 9:30 am to 6 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Dr. Talivaldis Smits, can be reached at (703) 306-3011. The facsimile phone

number for this Technology Center is (703) 305-9508.

Any inquiry of a general nature of relating to the status of this application should be directed to the Technology Center 2600 receptionist, whose telephone number is (703) 746-6055.

1/11/05

MCK

RICHEMOND DORVIL
SUPERVISORY PATENT EXAMINER